University of Groningen

Workload control in job shops, grasping the tap

Land, Martin Jaap

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2004

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Land, M. J. (2004). *Workload control in job shops, grasping the tap*. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-05-2023

Appendix A (article)

Published in Int. J. of Production Economics, 1996, Vol. 46-47 , pp 535-538 as:

Workload control concepts in job shops: A critical assessment

Martin Land, Gerard Gaalman

Abstract

In the case of production environments with job shop characteristics, much research has been done on partial control such as priority dispatching. The development of comprehensive control concepts lags behind. However, the principles of workload control (WLC) have been elaborated to more comprehensive production control concepts. WLC concepts buffer the shop floor against external dynamics by creating a pool of unreleased jobs. The use of workload norms should turn the queueing of orders on the shop floor into a stationary process which can be characterised by an equilibrium.

This paper compares and discusses the concepts of WLC. Assumptions of stationarity implied in the workload norms are exposed. A subdivision of workload definitions is chosen as a starting-point to trace assumptions of stationarity. The assumptions highlighted relate to the shop floor situation and make demands upon the job release function. An obvious conflict between timing and balancing within the job release function leads to an examination of stationarity requirements on the job pool contents.

The analysis of stationarity requirements within existing production control concepts provides guidelines for developing production control concepts for job shops working under dynamic circumstances.

A.1 Introduction

Traditionally, the job shop is a type of production environment which can be found in mechanical industry, particularly in component manufacturing. More recently, semiconductor fabrication has led to job shop situations. Job shops are characterised by a wide variety of products with variable routings and processing times. Job shops have a functional layout with universal equipment. Production takes place according to customer specification and in small batches.

Typical job shops have to work under very dynamic circumstances, both internally and externally. External dynamics relate for instance to rush orders, the product mix and volumes demanded, while internal dynamics may relate to machine breakdowns, production rates, operator absenteeism, quality problems, production yields, etc. We call a modelled job shop dynamic if the probability distributions, which describe the variables, are non-stationary and change in the course of time.

The bulk of literature on job shops has been devoted to priority dispatching. Surveys show hundreds of priority rules to be applied on the shop floor [Panwalkar & Iskander 1977, Blackstone et al. 1982, Ramasesh 1990]. Another research field receiving much attention is the assignment of due dates [e.g. Cheng & Gupta 1989]. Generally, research on priority dispatching and due date assignment does not consider comprehensive production control concepts, but isolates single elements of production control. The development of comprehensive control concepts still lags behind [Hendry & Kingsman 1989].

A starting point in the development of more comprehensive concepts has been the introduction of input/output control, first introduced by Wight [Wight 1970]. Since then input/output control has been extended to a class of hierarchical capacity-oriented production control concepts for job shops [Bertrand & Wortman 1981, Tatsiopoulos 1983, Bechte 1988]. These hierarchical concepts control workload, both at the level of order entry and the level of order release to the shop floor. The former level relates to all planned/accepted jobs, the latter relates to jobs on the shop floor. The control of workload on the shop floor creates a backlog/pool of orders waiting for release. The pool is claimed to buffer the shop floor against external dynamics. With this claim, the class of production control concepts using workload control might be attractive for use in job shops which are subject to dynamic circumstances.

This paper assesses how workload control (WLC) concepts deal with the dynamics of job shops. Comparing existing WLC concepts, we expose underlying assumptions of stationarity and corrections for violated stationarity assumptions. In order to compare the concepts we consider the classical job shop model, consisting of a set of work stations, each station concerning one specific capacity type, required for one specific operation on a job. We do not restrict ourselves to the pure job shop, the common model in most simulation studies. That means, capacities of work stations are not necessarily balanced, and job routings are not completely random with equal probability for each work station to be visited in each stage of job progress.

Section 2 elaborates the WLC paradigm. The analysis of three WLC concepts from the release point of view in Section 3 leaves us with three different workload definitions worth further investigation. It appears a useful starting point for our assessment in Section 4, as the definitions and the corresponding workload norms expose the stationarity assumptions relating to the shop floor situation. The release function must provide for the stationary workload on the shop floor. Section 5 discusses the obvious conflict between a timing and a balancing function of job release which lead to an exposure of stationarity requirements relating to the job pool. We summarise our analysis for the three referenced WLC concepts by means of a table.

A.2 The workload control paradigm

WLC conceptualises the job shop as a queueing system. In front of each work station, an arriving job finds a queue of jobs waiting to be processed. The principle of WLC concepts is to control the length of these queues. The main instrument for this purpose is the release decision. The release decision allows a job to enter the queue of its first work station in the shop. Once released, a job remains on the floor until all its operations have been completed. The progress of jobs on the shop floor is controlled by priority dispatching at each work station.

WLC concepts do not release jobs to the shop floor if they are expected to cause queue lengths to exceed certain workload norms. It results in a pool of jobs waiting for release. As illustrated by Figure A.1 we refer to waiting time in the pool as the *pool time* and to the interval between release and completion of a job as *the shop floor flow time*. The shop floor flow time of a job can be subdivided into *station flow times*. The pool is a new object of control. Unrestricted acceptance of jobs at the entry could cause excessive pool times.

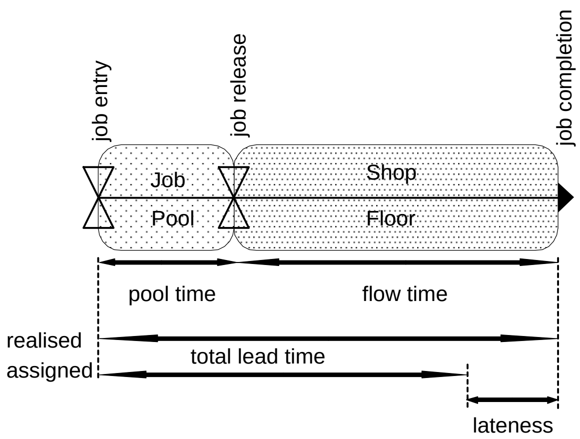

Figure A.1: Lead time components

A hierarchical control concept emerges [Kingsman et al. 1989], with three levels which respectively relate to job entry, job release and priority dispatching (Figure A.2). At each level, we distinguish two means of control, input control and output control. Input control regulates the allowance of jobs to the next stage, respectively accepting jobs for entry into the pool, releasing jobs to the shop floor, and dispatching jobs for processing (thus allowing a job to enter the queue of its next operation). On the output side, capacity management contributes to the control of workload through regulation of the outward flow, by means of respectively medium-term, short-term and daily capacity adjustments [e.g. Park & Bobrowski 1989]. In addition due date assignment or due date acceptance takes place at job entry. This paper concentrates on the input side.

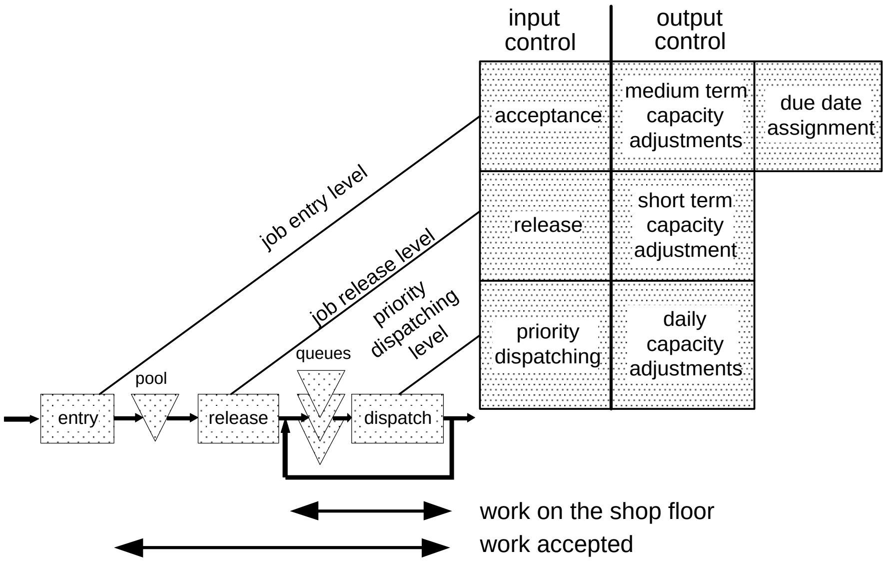

Figure A.2: The hierarchical WLC concept

The job entry level is very important, if one can influence the incoming orders. In that case, order acceptance and due date assignment/acceptance can support the release decision, providing it with a 'releasable' set of jobs, thus keeping pool times small. In fact, the job pool between entry and release acts as the visualised imbalance between job supply and production capacities.

The role of priority dispatching in WLC is a very modest one, because the choice among jobs is limited due to short queues. Generally, WLC concepts favour dispatching priorities such as first-come-first-served (FCFS) which stabilise operation flow times or due date oriented priorities which correct progress differences among jobs. These kinds of priorities facilitate a good timing of job release.

At the release level the use of workload norms controls the work station queues. The control of queue lengths, resulting in short and predictable flow times, is the key to both lead time and due date performance [Bertrand 1983]. However, the major strength of WLC concepts is withholding jobs from the shop floor, reducing average queue lengths. Besides a reduction of work-in-process, withholding jobs from shop floor has numerous additional advantages as it enables management to delay final production decisions [Irastorza & Deane 1974]. It reduces waste due to cancelled orders, facilitates later ordering of raw materials, takes away the need of expediting of rush orders, etc. Fluctuations in the incoming order stream should be absorbed by the pool. Altogether, it should create a stable stationary situation on the shop floor.

Only restricting queue lengths is generally not sufficient. If average queue lengths decrease but variances do not, the idle time at work stations will increase. This situation is not allowable for the common job shop, where many work stations can be temporary bottlenecks. The loads of potential bottlenecks should be kept close to a norm level instead of below a norm level. The release function which aims at short queue lengths and a reduced variability of queue lengths is called *load-balancing* within this paper.

Simulations of release rules with limited balancing qualities often show a deteriorated lead time performance. This has made the influence of 'controlled release' a topic of scientific research [Melnyk et al. 1991, Kanet 1988]. In practice, WLC concepts prove to have a positive effect on lead times [Wiendahl 1992], a result often attributed to improved 'shop floor transparency'.

In summary, WLC concepts try to create a situation on the shop floor of short and stable queues. A pool of unreleased jobs buffers the shop floor against external dynamics, the incoming non-stationary job stream. The queueing of jobs on the shop floor is turned into a stationary process. Release performs a key-role in reaching this stationary situation. It is the most elaborated function within WLC concepts. Therefore, we will compare and assess existing WLC concepts from the release point of view.

A.3 Existing WLC concepts

In the preceding section we have seen that release should control the queue lengths in front of each work station. The queues must be short and stable, the load-balancing function. On the other hand, each job should be released timely with respect to its planned due date and expected flow time, the *timing* function.

Leaving out capacity decisions at the release level, two components of the release decision are distinguished: a *sequencing* decision and a *selection* decision. The sequencing decision can be described as the setting of priorities for jobs to be released, 'selection' decides whether a job will be released or not at some specific moment. Most WLC concepts focus the sequencing decision on timely release and create due date based sequences. Taking into account this sequence, release selects a set of orders that keep the workload of work stations at certain norms. These workload norms are the main instrument of workload control.

For three WLC-concepts we discuss the release decision, the workload definition applied, and the determination of the corresponding workload norms. In addition, we present some developments which provide new ideas for release procedures within WLC.

A.3.1 Bechte's WLC concept

The release procedure proposed by Bechte [Wiendahl 1987, 1995, Bechte 1988, 1994] builds on three parameters: a *release period*, a *time limit* and a *load limit*. The decision to release jobs is taken periodically, at the beginning of each *release period*. All jobs in the pool are sequenced in order of their *planned release date*. The planned release date is determined by backward scheduling from the job due date: *norm station flow times* for all work stations in the routing of the job are subtracted from its due date. All jobs within the *time limit* from their planned release date are candidates for release. In the established sequence, jobs are released, until the workload norm of a work station, the *load limit*, is exceeded for the first time. All other candidates visiting this station have to wait in the job pool until the next moment of release. The selection process goes on for the remaining candidates.

The workload considered in the concept of Bechte is the queue length at a work station (in units of processing time). The workload is controlled by the load limit. The load limit LL_s of a work station *s* consists of two components: the planned output during the release period and the planned queue length at the end of the release period. The actual output O_s during the release period and the actual queue length Q^E_s at the end of the release period satisfy the balance equation:

$$Q^E_s + O_s = Q^B_s + I_s$$

with Q^B_s : the queue length at the beginning of the release period,
I_s : the input to the queue from jobs arriving during the release period.

The release decision at the beginning of the release period must bring $Q^E_s + O_s$ at the norm level LL_s. The above balance equation is used. Q^B_s is known at the moment of release, the queue input I_s is influenced by the jobs on the floor upstream of s and by the release of new jobs, see figure A.3. Some of them will arrive at s, some will not. Bechte estimates the input during the release period by means of the *load conversion algorithm*:

> If the workload of station *x* with a planned output component PO_x reaches its limit LL_x, a fraction PO_x/LL_x of the workload is planned to pass the station. Therefore, the probability that job j in the queue of *x* passes station *x* during the release period is estimated by PO_x/LL_x. The probability that job *j* reaches the queue of work station *s* is the probability that job *j* passes all its remaining upstream stations (the set U_{js}). This probability Pr_{js} is estimated by the product $\prod_{u \in U_{js}} (PO_u / LL_u)$. Suppose the processing time of job *j* at station *s* is p_{js}, then the expected input to the queue of station *s* from all upstream jobs (the set J_{us}) is estimated by $\sum_{j \in J_{js}} Pr_{js} \cdot p_{js}$.

First, load conversion is applied to estimate the input to the queue from jobs actually on the shop floor. Next, new jobs are released and their input is estimated until the estimated workload of a work station reaches its load limit. Notice that, within the load conversion procedure, the actual upstream positions of jobs at the time of release have been taken into account.

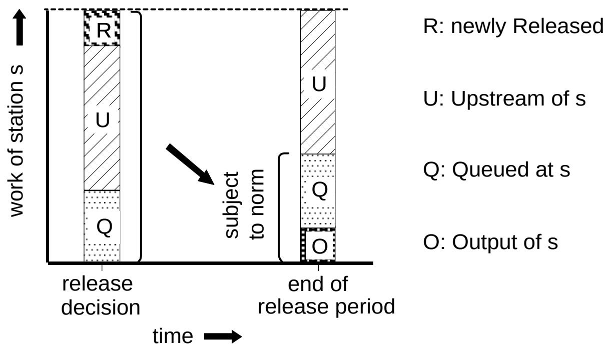

Figure A.3: Bechte: workload at the end of the release period estimated at the time of release

The workload norm LL_s is derived from the norm station flow time $NSFT_s$ of the work station. It assumes the following steady-state relationship [Wiendahl 1987, 1995]: $LL_s = PO_s + (PO_s/T)NSFT_s$,[10] with T: the length of the release period. The next step should be the determination of realistic norm station flow times, as they are essential elements of both workload norms and planned release dates. Nyhuis [Nyhuis 1992] presents a theoretical approach to estimate realistic norm values for this concept. Till now, trial-and-error determination, step-wise lowering norms, seems to be most successful for practical situations.

A.3.2 Bertrand's WLC concept

Bertrand developed a WLC concept for the diffusion department of a semiconductor plant [Bertrand & Wortmann 1981]. Bertrand does not discuss the release sequence, but elaborates the workload norms extensively. The release decision is taken periodically and the release of jobs is allowed if the workload of each work station remains below its norm value.

The workload considered in this WLC concept differs from the workload considered by Bechte. The workload definition of Bertrand covers the processing time of all jobs on the shop floor which still have to be processed at the work station concerned. The corresponding workload norm consist of two components: the planned work station output during the release period and the planned quantity of work upstream or in the queue at the end of the release period. An extended balance equation can be used to determine the actual workload of a work station *s* at the end of the release period:

$$(U^E_s + Q^E_s) + O_s = (U^B_s + Q^B_s) + R_s$$

with U^E_s : the processing time (on *s*) of jobs upstream at the end of the release period

U^B_s : the processing time of jobs upstream at the beginning of the release period

R_s : the processing time of jobs released at the beginning of the release period

All other variables as defined before.

At the moment of release the right-hand side of this equation is completely known. The processing times of all jobs which are newly released are the input to the

[10] Here $NFST_s$ is not determined as the norm station flow time for a job but for a unit of processing time.

workload. Thus, the release of new jobs directly influences the workload (see Figure A.4). The release decision can be made without a sophisticated estimation procedure.

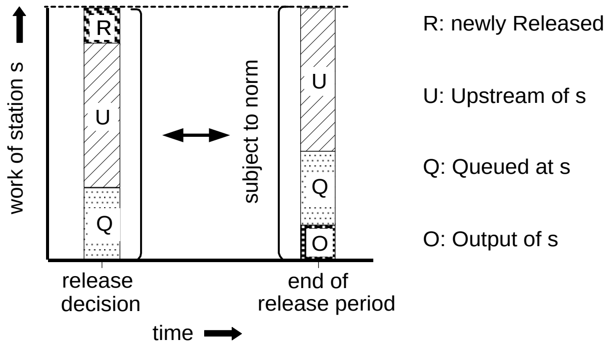

Figure A.4: Bertrand: the workload subjected to the norm is known exactly upon release

The determination of correct norm values is more complex for this workload definition. It is especially difficult to determine an accurate norm for the quantity upstream at the end of the release period. Bertrand calculates the norm as follows:

If the flow time of a job *j* at a work station *s* equals the norm station flow time NSFT_s, *j* will queue at this work station for NSFT_s time units. According to the workload definition, job *j* will be part of the workload of station *s* during its stay at stations upstream of *s* (the set U_{js}) as well. So, the norm pre-station flow time of job *j* is $\mathrm{NPFT}_{js} = \mathrm{NSFT}_{js} + \sum_{u \in U_{js}} \mathrm{NSFT}_{js}$. As long as job *j* is part of the workload, it increases the workload by its processing time p_{js}. In the course of time job *j* will contribute $\mathrm{NPFT}_{js} \cdot p_{js}$ to the cumulative workload of *s*. Now, Bertrand uses a set of jobs *J* which is supposed to be a good representation of the total population of jobs. All jobs of *J* together create a cumulative workload of $\sum_{j \in J} \mathrm{NPFT}_{js} \cdot p_{js}$. If the average output per release period of length T equals the planned output POs, it takes $\sum_{j \in J} p_{js} / \mathrm{PO}_s$ periods or $(\sum_{j \in J} p_{js} / \mathrm{PO}_s) \cdot T$ time units to process all the jobs of *J*. The planned average workload during this interval will be $(\mathrm{PO}_s / T) \cdot (\sum_{j \in J} \mathrm{NPFT}_{js} \cdot p_{js} / \sum_{j \in J} p_{js})$, which completes the calculation of the second norm component. It appears to be the product of the planned utilisation level and a weighted average of the norm pre-station flow times.

Finally, Bertrand adds the planned output component and sets the norm to

$$\mathrm{PO}_s + \frac{\mathrm{PO}_s}{T} \cdot \frac{\sum_{j \in J} \mathrm{NPFT}_{js} \cdot p_{js}}{\sum_{j \in J} p_{js}} .$$

In principle, this norm calculation applies to all work stations. For low-utilised stations a workload slightly higher than this norm is allowed if the actual job mix gives reason to it. Here, the effect on flow times will be small. Notice that the norm value calculated increases with the number of upstream stations. Roughly speaking, the norm value depends on the average work station position within *J*. The norm accounts for the work station position within a presupposed set of jobs. The release decision does not use information on the actual upstream positions of jobs at the moment of release. As for Bechte, the determination of realistic norm station flow times is open to question.

A.3.3 Tatsiopoulos' WLC concept

Tatsiopoulos [Tatsiopoulos 1983] developed a WLC concept for a small subcontracting component manufacturer. The concept has been elaborated by Kingsman and Hendry [e.g. Kingsman et al. 1989]. The concept formalises three ways of job release [Hendry and Kingsman 1991]. The common *push release* takes place periodically, *intermediate push release* can be forced by rush orders or orders with retarded material availability, and an *intermediate pull release* can be triggered from the floor when a foreman sees his station threatened by unplanned idleness. The periodic release decision considers the orders in the sequence of their planned latest release date. The calculation of the planned release dates is rough compared with Bechte. For each job the same norm shop floor flow time is subtracted from the job due date. The release of jobs is allowed unless a workload norm is exceeded, which applies to the intermediate pull releases as well. Additionally, a minimum workload is suggested. Unfortunately, both the use and the calculation of the minimum norm are not further elaborated.

Commonly, this WLC concept applies the same extended workload definition as the concept of Bertrand. We restrict ourselves to another workload definition, applied in the WLC system implemented by Tatsiopoulos [Tatsiopoulos 1983] and also mentioned in [Hendry & Kingsman 1988] and [Tatsiopoulos 1993]. This definition covers all work on the shop floor, even work completed at the work station concerned. For each work station a norm is set for the accumulated processing times of jobs upstream, job in the queue, and jobs downstream. The corresponding actual workload satisfies the following balance equation (see Figure A.5):

$(U^E_s + Q^E_s + D^E_s) + C_s = (U^B_s + Q^B_s + D^B_s) + R_s$

with D^E_s : the processing time (on *s*) of jobs downstream at the end of the period,

D^B_s : the processing time of jobs downstream at the beginning of the period,

C_s : the processing time of jobs which leave the shop during the release period.

All other variables as defined before.

Again all right-hand side components are known at the moment of release. The WLC concept does not clarify whether the shop output C_s from jobs fully completed during the release period is included in the workload norm. Notice that the workload definition further simplifies keeping up with the actual workload as it avoids the need for data regarding the completion of single operations. The completion of the job can be reported when it leaves the shop floor.

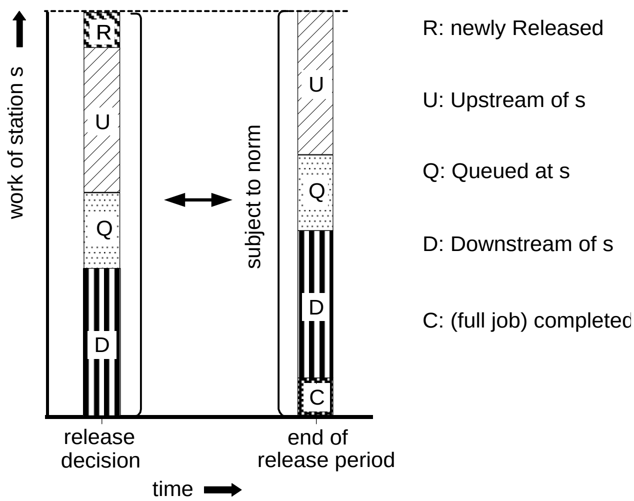

Figure A.5: Tatsiopoulos: a broad workload definition including downstream work

Hendry and Kingsman suggest that the workload definition enables the use of the same norm value for each work station [Hendry & Kingsman 1988]. This norm value is proportional to the maximum acceptable shop floor flow time.

A.3.4 Other methods of controlled release

Other ideas for release procedures within WLC are provided by [Glassey & Resende 1988] and research of Wein [Wein 1990, Wein 1992, Wein & Chevalier, 1992]. Both suppose continuous release opportunities. As a consequence, job release takes place whenever a workload falls below its norm, instead of periodic replenishments. Both studies assume explicitly that all random variables are stationary.

The *starvation avoidance* policy of Glassey and Resende focuses at the avoidance of idle time at a bottleneck station. The policy is only elaborated for the very simplified situation of a flow shop with only one job type and one bottleneck station. However, their workload definition is an interesting contribution to the spectrum we recognised. It includes the processing time of a job in the workload of the bottleneck station, when the job's remaining processing time upstream is below a *critical time factor*. Thus, jobs in the queue and part of the jobs upstream of the work station are included.

Wein applies the same workload definition as Bertrand, the accumulated processing times of jobs upstream and in the queue of the work station. The difference between the concept of Wein and the previous ones can be found in the release sequence and the use of norms. The release procedure is elaborated for a situation with two work stations. Wein combines norms for the absolute workloads of both stations with norms for the ratio between the workloads. Figure A.6 graphically depicts the workload conditions which require new releases. The shape of the area requiring new releases differs from the common rectangle area which results from absolute norms. The cut resulting from the ratio norms represents the principle that a better ratio between the workloads allows for lower workloads. Wein primarily sequences the jobs in order of their due date. But, when the difference between two due dates is below a certain limit, priority is given to the job which best restores the workload ratio between the stations, that is the job which contributes most to the smallest workload. Thus, the control policy manipulates the shop situation in the direction of the internal corner *c* of the shaded region in Figure A.6. Point *c* represents the smallest combination of workloads required.

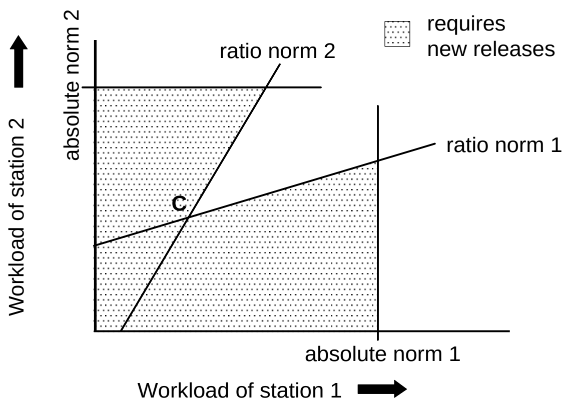

Figure A.6: Workload conditions requiring release in the concept of Wein

A.4 Workload definitions and shop floor stationarity

The discussion of WLC concepts highlights both differences and similarities between WLC concepts. Especially the use of different workload definitions and corresponding norms is worth further investigation. All WLC concepts use norms for the quantity of work allowed on the shop floor. As jobs are released periodically, norms are set for the desired situation at the end of the release period. More precisely, norms are set for each work station on the shop floor, as the WLC concepts aim to control the queue length in front of each work station. Though principally the objective is to control the load in the queue of each station, we observe the use of extended workload norms. Bertrand includes the work content of work upstream and Tatsiopoulos all work on the shop floor, both upstream and downstream (see Figure A.7).

The reasons for extensions stem from practical perspectives. Our discussion of the Bechte concept highlights that restricting the workload to the work station queue requires estimations of the input to the queue. Since all jobs upstream are candidates, a workload that includes all upstream work eliminates the need of input estimations. Upon release, a job contributes immediately to the workloads. If downstream work is also incorporated, the workload of each station will only change at job releases and job completions, and the release decision no longer requires a record of operations completed at each work station.

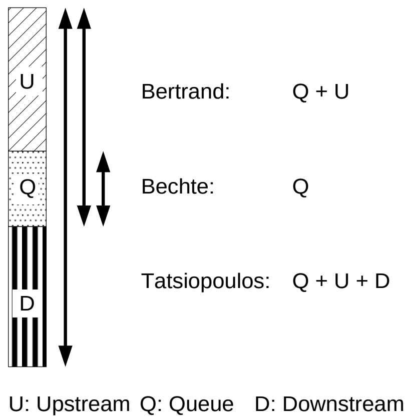

Figure A.7: The subdivision of workload definitions

If an extended workload norm should control the load in a queue during the release period, assumptions will be inevitable. For each type of workload addressed by the workload norm we expose the underlying assumptions:

A.4.1 Bechte: Queue only

In general, the contents of the work station queues are not directly influenced by the release of jobs. Jobs may have to pass other (upstream) stations first. Since the arrival of jobs is influenced by many uncertain factors, simplifying assumptions are necessary to obtain a simple estimation procedure. The WLC concept of Bechte accounts for the actual upstream positions of jobs at the time of release. So contrary to the other concepts, this concept does not make assumptions about these positions. The assumptions of Bechte are restricted to the flow of jobs during the release period as it is estimated by load conversion, and to the actual volume of the workloads upon release. The load conversion procedure evoked a number of criticisms, criticisms for the larger part published in German literature [Adam 1988, Adam 1989, Häfner 1992, Hansman 1993, Knolmayer 1991, Greiner 1989]. Without going into detail, we might say that most of these criticisms relate to the assumption of unrestricted divisibility of the workload:

(1) The estimated probability PO_x/LL_x that a job passes a station *x* neglects the fact that each job as a whole must pass the station during the release period and not a fraction of its processing time. If its processing time is large, the actual probability will decrease.

(2) The product form of Pr_{js} suggests that the probability to pass a station is independent of the number of stations already passed during the release period, though each operation will delay a job, at least for its processing time. The probability that a specific job reaches a station might be estimated more accurately by considering its planned pre-station flow times in relation on the length of the release period.

If the workloads are large relative to processing times, the estimations show increased accuracy. Ambiguously, the paradigm of workload control forces the workload in the opposite direction. Since it aims at small workloads, workload control will lead to a more restricted divisibility. We observe other assumptions which relate to the volume of the actual workload implied in the estimated probability PO_x/LL_x:

(1) The actual output O_x of the supplying stations should equal the planned output PO_x.

(2) The actual workloads estimated from $Q^B_x + I_x$ should equal the norm LL_x.

Though the assumptions may seem obvious, the fact that they should hold for each work station, imposes strong requirements to the set of jobs on the floor.

A.4.2 Bertrand: Queue and upstream included

It has been shown that the inclusion of upstream work in the workload norm avoids the need of predicting queue inputs at the moment of release. However, the easier determination of the actual workload at the moment of release rebounds upon a more complicated determination of the workload norm. It is possible to disaggregate Bertrand's second norm component into an element for the load upstream and an element for the load in the queue. The actual workload elements U^E_s and Q^E_s should correspond with their norm parts. Otherwise, the queue might be idle while the workload is at its norm level. So, a stable composition with respect to the shares of the upstream and the queue elements of the workload is assumed. Bertrand does not check this assumption at the moment of release.

These general considerations about the workload composition can be elaborated in more detail. The calculation of the workload norm reveals the detailed assumptions. The norm is calculated for a specific set of jobs *J* which is supposed to be representative for the future portfolio of jobs. One might wonder what happens if the portfolio on the shop floor is going to differ from *J*. The calculation shows that the workload norm of a station *s* increases with the number of (upstream) stations to be visited before *s*. If the jobs on the shop floor visit *s* later on average than the jobs of *J* do, the actual load upstream should exceed its norm component in order to provide *s* with the planned queue load. The actual upstream positions of jobs within the workload of *s* are not checked. Thus, one must assume that the relative position of each work station within the actual job mix on the floor varies little, and that on average it equals its position within the presupposed set of jobs *J*. More exactly, the assumption relates to the pre-station flow time characteristics of the jobs on the floor. The actual mix of pre-station flow times weighted by the processing times of jobs should be stationary, as the calculated norm component

$$\frac{\mathrm{PO}_s}{T} \cdot \frac{\sum_{j \in J} \mathrm{NPFT}_{js} \cdot p_{js}}{\sum_{j \in J} p_{js}} \quad \text{shows.}$$

Bertrand corrects the calculated norms. The correction allows for exceeding the norms of low-utilised stations. It provides room for deviations from the norm mix *J*. However, we note that the corrections only seem suitable for small deviations around the means. They will not be adequate for a shift of the mean or heavy incidental disturbances.

The idea of Glassey and Resende to control the quantity of work within some time distance from a work station adopts a middle course between the workload definitions of Bechte and Bertrand. As within the concept of Bertrand, the workload is an aggregate of the actual queue contents and future contents, and no estimation of the queue input is made. In contrast with Bertrand, Glassey and Resende account for the upstream position for jobs at the time of release, but in a less detailed way then Bechte. Since the release of a job only affects the workload of its first work stations directly, one should assume that the set of jobs released provide their downstream stations with a stable future load. The non-periodic release decision enables a fast reaction to load deviations.

A.4.3 Tatsiopoulos: Queue, upstream and downstream included

Control of the work upstream of a work station *s* bears importance for the control of the work station queue, as the work upstream incorporates the future queue load of station *s*. Including work downstream of *s* in its workload norm does not contribute to the control of the queue. The jobs concerned already passed the queue of *s*. Any fluctuations of the actual downstream component will needlessly influence the release decision. An actual downstream component exceeding its corresponding norm component causes the decision to slow down the release of jobs. As a consequence, the work station considered may get idle even when there is plenty of work to be released. Thus, one should assume that on average the actual downstream component equals the norm component and that its fluctuations are be limited.

Violations of this assumption may have serious consequences. Since the downstream component will be relatively large for a station performing preparatory operations (i.e. a gateway station), these stations are particularly threatened by (unnecessary) idleness. Tatsiopoulos obviates this problem by providing gateway stations with the possibility of pull-release.

The workload norm no longer depends on the average position of a work station in the mix of jobs. No assumptions have to be made regarding the work station position in the actual set of jobs on the floor. A shift of the average work station position will have no effect in the long term.

In general, the WLC concepts must realise a stable input to the work station queue in order to control its length. Independent of the type of workload norm, it is in general not possible to influence this input directly by the release of jobs. One depends on assumptions regarding the input to the queue. First, the output rate of supplying stations is assumed stable. Stationarity of work station capacities is a prerequisite for stable output rates. Second, the mix of jobs released must have stationary characteristics. Bechte accounts for the actual upstream positions of jobs to estimate the input during the release period and his assumptions are restricted to the load conversion estimation procedure. Bertrand and Tatsiopoulos do not check the actual positions of jobs at the beginning of the release period. They assume that the extended workload which is subjected to their norms provides the planned input to the queue. These assumptions will be violated if the characteristics of the actual workload on the floor differ from the characteristics supposed within the norm calculation. All WLC concepts use a relationship between workloads and planned flow times. These relationships only hold in a stationary situation and with the assumption of all work stations loaded up to their workload norm. In summary, we may say that stationarity of characteristics is assumed for both jobs and capacity on the shop floor.

Realising the stationarity of capacity characteristics goes beyond the span of control of WLC concepts, realising stationarity of characteristics for the jobs within the workload is claimed to fall within. Till now we did not address the question whether, even if the characteristics of the actual workload have been checked, a release policy will be able to provide the work stations with the required stationary workload. The next section assesses the ability of the release policies to create the required workload.

A.5 The timing/balancing conflict and pool stationarity

The preceding section pointed out that all WLC-concept make assumptions which relate to mix of jobs on the shop floor. Since the release decision should provide the shop floor with this mix, the assumptions impose requirements on the release decision. A minimum requirement within all WLC-concepts is that the volumes of the actual workloads upon release equals the workload norms.

Stable workloads, equal to norm values, should be guaranteed by the load-balancing function of release. Only if load-balancing functions well, the queues of work stations will be stable. Stable queues should keep flow times at their planned level. Planned flow times in turn determine the planned release date of a job. So, a precise timing of the release moment of a job depends on stable flow times. As a consequence, this timing-function of release depends on an effective load-balancing function to realise a good due date performance. The question is whether a good timing of job release also allows for sufficient load balancing. By assessing the release procedures of the WLC concepts, we will determine under which conditions the load-balancing and timing function co-operate.

We argue that the referenced concepts deal with order release in a one-sided way. Accurate timing is provided by the sequence in which jobs are considered for release. But, the load-balancing qualities are limited. The release procedures fit jobs into the workload in the predetermined sequence of planned release dates. Once a job fits, its release will not be reconsidered. This can be seen as a greedy algorithm. As a result, some workloads might be far below their norm, because the workload of one station reaches its norm. The release sequence could have been reconsidered in order to approximate the complete set of norm values more closely. In particular if the accepted order portfolio requires high utilisation levels, WLC-concepts may require better balancing properties to create sufficient throughput capability [Land & Gaalman 1994]. An example of a completely balance-oriented approach is presented by Shimoyashiro et al. [1984].

The release policy of Wein can be a first step in the development of more powerful release policies. It shows better balancing properties than the release policies of the referenced WLC concepts and carefully weighs balance requirements against job due dates. The policy does not require each station to be loaded up to a fixed norm. Instead it allows small fluctuations of the ratio between workloads, but the better this ratio, the smaller the volumes of workload required.

The job pool makes the balancing function less sensitive to the dynamics of the incoming order portfolio. A larger pool increases the choice of jobs to fill workload gaps. That way, the capacity requirements of the incoming stream of jobs are smoothed by the pool. It depends on the size of the pool to what extent fluctuations can be absorbed. However, a larger pool increases pool times and deteriorates lead time performance. Thus, lead time requirements restrict the size of the pool. This restriction may create a conflict between the load-balancing and timing function of release. At a certain moment jobs require release according to their planned release date. If the set of jobs requiring release do not fit into the workload norms, jobs will be delayed until the next moment of release and due date performance will deteriorate. Only if the load contribution of the job set requiring release does not show excessive peaks for any work station, conflicts between the timing and balancing functions can be avoided. So, a certain stationarity of the job pool contents must be required. Melnyk et al. [1992] discern the same problem. Their simulation results indicate a more effective release, when release is preceded by smoothing of the workload.

Only small fluctuations around stable means can be absorbed by the pool. Strong dynamics (unstable means, etc.) related to the incoming order portfolio will not create sufficient stationarity within the job pool. Existing WLC-concepts confronted with strong dynamics of the incoming stream of orders will depend on either high flexibility of capacity or possibilities to reject stationarity disturbing orders at the entry level. Till now, output control and order acceptance have been the least elaborated elements of the WLC concepts. An exception should be made for recent research on order acceptance by Hendry and Kingsman [1993].

Of course, norm values can be adjusted continuously in dynamic situations. With the help of linear programming techniques, Zäpfel and Missbauer [1993] determine new norm values, whenever dynamics of the incoming order stream give rise to this. Even if adequate determination of optimal norm values is possible, this will lead to cumbersome and nervous procedures for job shops which are exposed to strong dynamics.

During a short time interval, we might assume a stationary situation. Even then, it is still questionable whether the actual workloads must be exactly adjusted to a norm value upon release. Also in a stationary situation, workloads fluctuate without deteriorating performance: reacting to each deviation from the norm might lead to over correction. Instead, we might release constant quantities of work and only correct these quantities for fluctuations that exceed some 'normal-variance-based' bounds. Such bounds may be able to handle an increased range of dynamic fluctuations without causing over correction, as the norm adjustments of Bertrand aim at reducing over correction of small load fluctuations. The above approach has proven its value in the field of statistical quality control.

A.6 Conclusions and suggestions for further research

WLC concepts buffer the shop floor against external dynamics by creating a pool of unreleased jobs. The use of workload norms should turn the queueing of jobs on the shop floor into a stationary process. Here, the release decision performs a key-role. WLC concepts translate the term ’control’ to ’maintenance of workload norm levels’.

However, each type of workload norm brings about a series of stationarity assumptions. Roughly speaking, WLC concepts assume stationarity of the shop floor situation. They depend on a certain stationarity of the job pool contents to create this stationary situation. Otherwise, the release decision will be confronted with conflicts between its load-balancing and its timing function.

Though a large pool buffer may protect the shop floor against external dynamics, it puts high pressure on lead times. Consequently, the WLC concepts correct for violations of internal stationarity assumptions, adjusting norms before release, or afterwards with intermediate releases. Table A.1 summarises the different workload norms, the assumptions and the formalised corrections.

The question arises whether all stationarity assumptions are necessary. Might it be possible to incorporate the reactions to dynamics in the frame of the control concepts? Continuous adjustment of norm values is a cumbersome procedure, since even the determination of accurate norm values is a complex decision, not yet crystallised. The many job shops exposed to strong dynamic circumstances require control concepts that handle dynamics in a more natural way. This provides an interesting domain for further research.

Even, if we suppose temporary stationarity, the existing WLC concepts, with their continuously changing release quantities, neglect the normal variability of stationary characteristics. Statistical quality control has embraced control concepts that only react to excessive variability or shifting means, the real out-of-control situations. Statistical production control concepts like workload control, might gain applicability by adopting this approach.

	Bechte	*Bertrand*	*Tatsiopoulos*
Workload subjected to norm (for release period T)	Station output during T + queue at the end of T	Station output during T + queue + upstream work at the end of T	Shop floor output during T + queue + upstream work + downstream work at the end of T
feedback frequency	Each completed operation	Each completed operation	Each completed job
norm determination	Directly derived from norm station flow times	Derived from presupposed job mix and their norm pre-station flow times	Equal to maximum norm shop floor flow time
upstream position of jobs versus work station position	Actual upstream positions of jobs used for estimation of queue input; workload norm independent of work station position	Actual position not checked; workload norm depends on position of work station within job mix	Actual positions unknown; workload norm independent of work station position
assumptions	- Each station loaded up to its norm - Simplifying assumption to estimate queue inputs	- Each station loaded up to its norm - Smooth queue input from upstream jobs - Actual job mix corresponds with pre-supposed mix with respect to pre-station flow times	- Each station loaded up to its norm - Smooth queue input from upstream jobs - Stable downstream workload component
corrections	No formalised corrections	Norm adjustment allowing small load fluctuations for low-utilised stations	Intermediate pull-release for idling work stations

Table A.1: Analysis of the concepts of Bechte, Bertrand and Tatsiopoulos

Appendix B (article)

Published in Production Planning & Control, 2002, Vol. 13 , No.7 pp 625-638 as:

The workload control concept: theory and practical extensions of Load Oriented Order Release

Jan-Wilhelm Breithaupt, Martin Land, Peter Nyhuis

Abstract

Workload control (WLC) has been elaborated in the early 1980s to a hierarchical production control concept for job shop manufacturing. In the 1990s the analytical research at the University of Groningen focused on assessing the strengths and weaknesses of this concept and on developing alternatives. Research at the University of Hannover focused on improving the practical applicability of the WLC concept. While the practical experiences confirm the strengths of the WLC concept, some extensions of the basic concept have been developed which may overcome some weaknesses suggested in literature. This paper aims at bringing the theoretical and practical knowledge regarding the WLC concept together. It gives a review and classification of strengths and weaknesses reported from analytical research and it discusses the extensions that have been developed based on practical experiences.

B.1 Introduction

Make-to-order companies and particularly those characterised by job shop production traditionally emphasise the importance of workload control. There, lack of workload control may lead to nontransparent shop floor situations with unreliable lead times and much expediting. In the early 1980s the principles of workload control were elaborated to a production control concept for job shop production. Three roughly comparable control concepts were developed in respectively Eindhoven [Bertrand & Wortmann 1981], Hannover [e.g. Bechte 1988] and Lancaster [e.g. Kingsman et al. 1989]. In fact, each suggests a hierarchical approach with three control levels relating to phases in the order flow. Figure B.1 shows how control of the accepted amount of work takes place at the order entry level, while control of the workload on the shop floor takes place at the release level. Priority dispatching remains for correcting progress disturbances among orders at the shop floor. At each level the decision must be made which orders can be allowed to proceed to the next stage (input control) and whether this requires capacity adjustments (output control). Since the early 1980s the order entry level has received particular attention in Lancaster [Hendry & Kingsman 1993], while researchers in Hannover elaborated several production control elements, such as load oriented lot sizing [Nyhuis 1991].

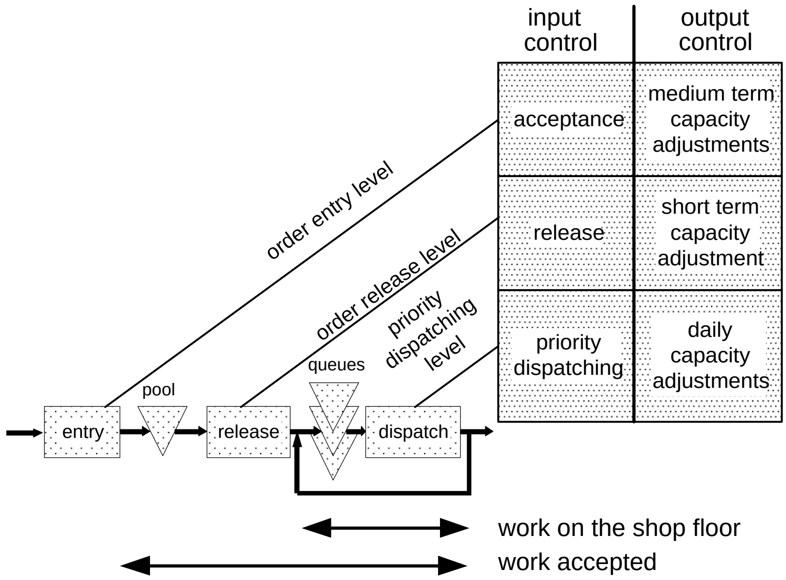

Figure B.1: The general framework of the workload control concept

In the late 1980s, a discussion started on the value of controlled release [Melnyk & Ragatz 1989, Kanet 1988]. Though job shop practice confirmed the value of controlled release, simulation studies showed adverse results. However, simulation studies generally focused on release methods that are less advanced than the methods used in the class of hierarchical concepts discussed above. These advanced methods are based on simultaneous control of the work released for several work centres, not only restricting but also balancing workloads. A further classification of release methods is given in [Bergamaschi et al. 1997]. Elaborating this classification, the performance of some advanced methods has been compared in [Cigolini et al. 1998]. Until the 1990s, studies analysing advanced release methods were mainly published in German literature [e.g. Adam 1988, Greiner 1989, Häfner 1992]. Most criticisms in German literature deal with the particular calculations in the release method developed in Hannover. Land and Gaalman [Land & Gaalman 1996a, 1998, Oosterman et al. 2000] compared and assessed the class of advanced release methods from a more conceptual point of view. Several strengths and weaknesses can be deduced from the analytic studies that have been performed.

The first implementations of the release method developed in Hannover (indicated as LOOR – Load Oriented Order Release) in industry are discussed in [Wiendahl 1991, 1993]. By now, LOOR has been applied in a number of software packages. In a market analysis of 210 PP&C- and ERP-systems [Fandel et al. 1998] an overview of implemented order release methods is given (figure B.2). Examples of well-known software-packages containing LOOR are SAP R/2 (SAP), debis-PPS (debis SHE), MAS 90 BWR (IBM), PIUSS PENTA and PIUSS-O (PSI).

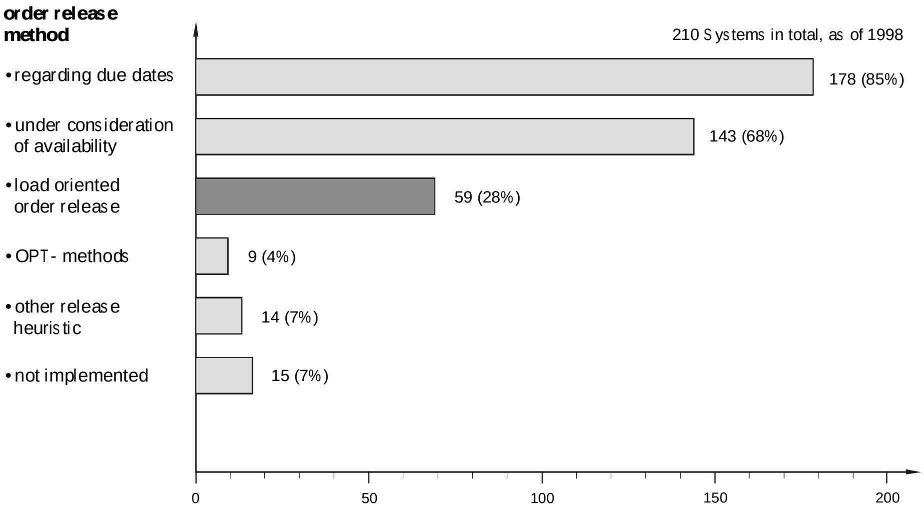

Figure B.2: Implemented methods of order release in PP&C- and ERP-systems [Fandel et al. 1998]

The use of the LOOR method in industry has led to some extensions and adjustments that have not been reported in literature yet. This paper discusses these extensions and adjustments, which may overcome certain weaknesses that have been suggested in analytic studies. The next section starts with a discussion of order release within the workload control concept and the LOOR method in particular. Next, a brief review of the strengths and weaknesses suggested in literature is given. Finally, the contribution of three extensions based on experiences with the LOOR method is discussed.

B.2 The workload control (WLC) concept

B.2.1 General principles

The workload control (WLC) concept recognises that job shop production inevitably shows queues of orders that compete for the capacity of each work centre. The WLC concept tries to create small and stable queues or, more precisely, low and stable levels of direct load. The direct load of a work centre is defined as the quantity of work resulting from waiting orders together with that of the order being processed.

The WLC concept smoothes the flow between the work centres by trying to release the right order at the right time. The complexity of this kind of input control results from the routing variety in job shops. After the release of an order, other operations may have to be completed before an order can be processed at a certain work centre. Thus, orders giving input to the direct load of a work centre may come either directly from release or indirectly from any other work centre. Several approaches have been suggested to smooth these combined inputs to the direct load.

The LOOR approach developed in Hannover is the most straightforward in controlling the direct load. It records the actual direct load of the work centres at the time of release and estimates the input to the direct load during the following planning period, using an estimation method called load conversion. The converted load, being the sum of the direct load and the estimated input, is kept at a norm level for each work centre by releasing the right amount of work. LOOR was first presented in the dissertation of Bechte [Bechte 1980] and extended by several authors [e.g. Wiendahl 1995].

Another approach is presented in the dissertation of Bertrand and Wortmann [Bertrand & Wortmann 1981] and in the dissertation of Tatsiopoulos [Tatsiopoulos 1983], later extended by Hendry and Kingsman [Hendry 1989, Hendry and Kingsman 1991]. Their approach is to aggregate the direct load and the indirect load of a work centre. The indirect or upstream load of the work centre is defined as the quantity of (future) work coming from orders that queue at other work centres for preceding operations to be completed. The sum of its direct and indirect load is called the aggregate load of the work centre. The release methods of this second approach use norms for the aggregate loads, instead of estimating the inputs to the direct load.

A third approach [e.g. Oosterman et al. 2000] combines elements of the two other approaches, using norms for adjusted aggregate loads. The underlying idea is that the aggregate load of a work centre should be corrected for the variable position of this work centre within the routings of released orders. The applied correction is such that the adjusted aggregate load of a work centre can be seen as an estimate of its future direct load.

All approaches use workload norms to create a smooth flow of work, but differ with respect to the type of load that is bounded by norms. More specifically, the contribution of orders to the load is different. With LOOR the contribution depends on the upstream distance of the order to the work centre considered. If norms for aggregate loads are applied, one accounts for the full operation processing times as soon as an order is released, irrespective of the number of operations to be completed before the work centre is reached. The adjusted aggregate load includes just a fraction of the operation processing time in case of downstream operations. Contrary to LOOR, the size of the fraction is not altered during upstream progress of the order; the depreciation of the operation processing time is proportional to the work centre position. Figure B.3 sketches the contribution of a job *j* to each type of load account in the course of time. The account concerns work centre *c* that performs the n_{jc}-th operation (in this case the third operation) of the job, with operation processing time t_{jc}.

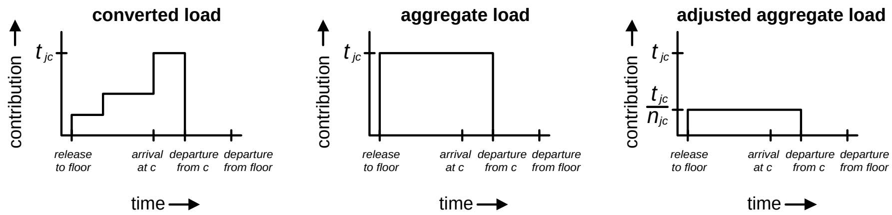

Figure B.3: Contribution of an order to the load calculation in each of the approaches

Besides controlling the workload, the release methods look at the relative urgency of orders. This is realised by considering orders for release in sequence of a planned release date. As the controlled direct loads should result in relatively constant work centre lead times, the planned release dates can be determined relatively easy. The next section will discuss the LOOR approach in more detail.

B.2.2 LOOR (Load-Oriented Order Release)

Figure B.4 overviews the release procedure of LOOR. This procedure, performed at the beginning of each planning period, requires a number of parameters to be predetermined: the length of the planning period, work centre lead-time allowances, an anticipation horizon, and loading percentages.

The first step in the release process is the backward scheduling of all issued shop orders that have not been released yet. It results in a list of urgent orders, sorted by planned release date. First, the planned release date for each order is determined as its due date minus all relevant work centre lead-time allowances. Next, all orders with a planned release date falling within the anticipation horizon (specified as a multiple of the planning period length) are classified as urgent. Only urgent orders will be considered for release in the following steps.

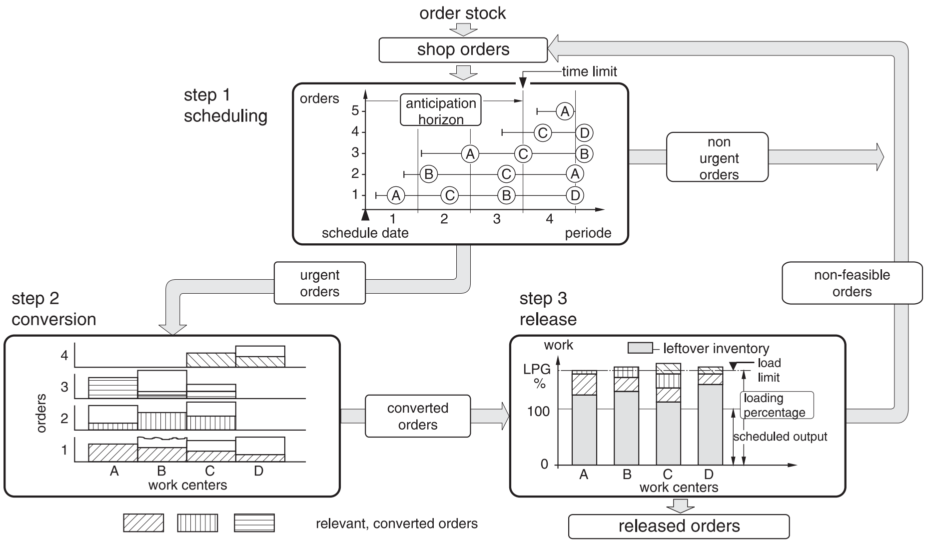

Figure B.4: Steps of Load-Oriented Order Release [Wiendahl 1995]

The second step determines the load contribution of the urgent orders by means of a procedure indicated as *conversion.* The conversion procedure accounts for the planned inventory conditions a job will meet, which depend on the level of the workload norms. Within LOOR, the workload norm of a work centre c is expressed as a percentage (>100%) of its planned output. It is called the *loading percentage* (LPG). The first operation of an order will be loaded onto the account of the corresponding work centre with its full operation processing time. To load a downstream operation of the order, the probability that the order will reach the corresponding work centre within the planning period is considered. The estimated probability ($POUT_p$) for the p-th operation to be completed within the next period is given by equation (B.1).

(B.1) $$POUT_p = PINP_{(p+1)} = PINP_p \times \frac{100}{LPG_{c(p)}}$$

where:

$POUT_p$	:	Estimated output probability for the p-th operation
$PINP_p$	:	Estimated input probability for p-th operation
$LPG_{c(p)}$	:	Loading percentage of work centre c, which performs the p-th operation

By means of equation (B.2) the probability that an order waiting for operation 1 will reach the downstream work centre that performs the p-th operation in the subsequent period is estimated.

(B.2) $$PINP_p = POUT_1 \times POUT_2 \times ... \times POUT_{(p-1)}$$

It is called the conversion factor CF_p. Based on equations (B.1) and (B.2), CF_p is calculated as follows:

(B.3) $$CF_p = \frac{100}{LPG_{c(1)}} \times \frac{100}{LPG_{c(2)}} \times ... \times \frac{100}{LPG_{c(p-1)}}$$

If all work centres have the same loading percentage, the equation is simplified to:

(B.4) $$CF_p = \left(\frac{100}{LPG}\right)^{p-1}$$

Finally, the contribution to a load account is determined for each operation by multiplying the operation processing time with the individual conversion factor CF_p.

In the third step, the urgent orders are successively considered for release. Starting with the order with the nearest planned release date, the converted operation processing times are loaded by trial onto the load accounts of the respective work centres. If none of the accounts of work centres required for processing of this order is blocked, it is released and loaded finally with its converted processing times onto the accounts of the corresponding work centres. As soon as the workload norm of an account is exceeded for the first time, this account will be blocked. Now the second order is tested in the same manner, followed by the third, fourth, etc.. If any operation of an order should be loaded to a blocked account, the entire order is entered into the list of non-feasible orders. Together with the non-urgent and possible rescheduled orders it is re-entered into the release procedure at the beginning of the next planning period.

The initial load account level, indicated as ‘leftover inventory’ in figure B.4, contains the load resulting from previously released orders. Previously released orders that are still upstream of the work centre have been converted similar to new orders, but the output probabilities of completed preceding operations have been excluded from the factor CF_p.

The above procedure, executed periodically, determines the set of orders that is released, considering both the inventory conditions on the shop floor and the urgency of orders.

B.3 Strengths and weaknesses

Analytic and simulation studies have revealed several strengths and weaknesses, some relating to the basic concept of order release within WLC, others being more specific for the LOOR method. The next section briefly reviews the strengths and weaknesses reported in literature. The first subsection discusses aspects that relate to the basic concept of using workload norms. Next, the qualities of the WLC methods in reducing and balancing workloads are assessed and finally the LOOR-specific strengths and weaknesses are analysed. The main elements are marked boldly and will be summarized in section 5.

B.3.1 Use of norms

An important and obvious quality of the WLC concept is that **it buffers the shop floor against the dynamics of the order flow**. If orders were directly released to the floor, capacity bottlenecks and conflicts regarding order urgency should be solved on the shop floor. With WLC, the shop floor situation is kept within norms, while the pool of orders waiting for release can be used to visualise possible problems in an early phase [Bechte, 1994]. It facilitates taking adequate measures such as capacity changes and due date adjustments in the order acceptance stage (see figure B.1). In fact, dynamics should be absorbed in the pool, while the shop floor is kept in a certain stationary state.

Many control approaches for job shop production risk the occurrence of a vicious cycle, resulting in an on-going increase of shop floor lead times. The **use of fixed workload norms in WLC avoids this so-called lead time syndrome**. [Plossl 1988] describes the vicious cycle which occurs if planned lead times automatically increase as loads increase:

1) Increasing planned lead times generates more orders for release immediately,
2) earlier release of these orders increases work centre loads, thus queue times lengthen,
3) actual lead times get longer also and more delivery dates are missed …

and the cycle repeats (Figure B.5). In fact this cycle occurs if a release method advances the release of orders when workloads increase. WLC breaks the vicious cycle in step 2, because release is restricted by the workload norms.

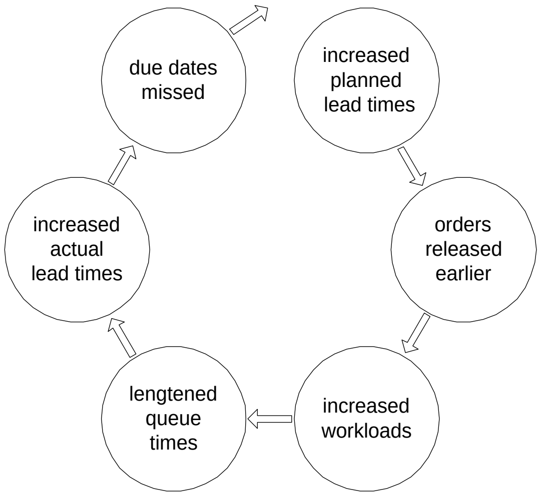

Figure B.5: The vicious cycle of production control [based on Plossl 1988]

The workload norms provide a convenient means to communicate at the interface between overall planning and shop floor control. If the work released complies with the agreed workload norms, overall planning may rely on shop floor lead times. Thus, the parameters create clarity between the parties involved in planning and control. Nevertheless, setting parameters may be delicate matter. In the case of LOOR, workload norms for each work centre, planned output, the anticipation horizon, and the length of the planning period have to be determined. Setting all these parameters must be done carefully. At low levels of work-in-process, performance appears to be **very sensitive to small changes of workload norms** [Land & Gaalman 1998]. Also **setting the anticipation horizon is delicate**. If the parameter strongly restricts the set of orders that can be considered for release, less balance of the workloads can be reached, and output will drop. Contrarily, a long anticipation horizon may result in a strong dispersion of due date deviations. For LOOR, the simulation results in [Perona & Portioli 1998] indicate that **performance is sensitive to the chosen length of the planning period**. Hendry et al. [1998] observe similar sensitivities in case of aggregate load norms.

It is shown in [Land & Gaalman 1996a] that each norm brings about a series of assumptions regarding stationary characteristics of the order mix. To check whether assumed conditions still hold, **WLC will require continuous monitoring and adjustment of workload norms and other parameters.** The importance of continuous adjustment in a strongly dynamic environment is illustrated in [Zäpfel & Missbauer 1993] and [Perona & Portioli 1996]. The SLAR method [Land & Gaalman

1998] has been developed to provide a starting-point for control of workloads without the need to determine norms.

B.3.2 Reducing and balancing workloads

The core of the WLC concept is the control of workloads. A WLC concept derives much of its strength from these controlled workloads. One obvious aspect is **that work-in-process is reduced**. In turn, the reduction of work-in-process leads to a **more transparent shop floor with less expediting** [Melnyk & Ragatz 1989, Wiendahl 1991], while the lean shop floor also **diminishes the dependence on sophisticated priority rules** [Bechte 1988]. As a drawback, there will be **fewer opportunities to reduce sequence dependent set-up time**s by choosing an efficient processing sequence on the shop floor.

If workloads were just reduced, the idle time of capacities could increase. Therefore, workload control does not just face the task to reduce workloads but also to balance workloads. Here, workload norms function as threshold values. Orders are selected for release, so as to approach the norms as well as possible. If an order does not fit into the workload norms due to a blocked account, the procedure will enable the selection of a less urgent order that does fit. Such an order improves the balance by filling the gap between the recorded workload levels and the workload norms. It creates a steadier load for the work centres. The steady direct loads prevent the shop from increased idleness, and furthermore help to **create lead time predictability** [Land & Gaalman 1996a]. With predictable lead times for each work centre, the improved timing of order release can in turn contribute to due date reliability.

Taking a different perspective, the influence of controlled release can be viewed in terms of order waiting times. Controlled release turns waiting time on the floor into waiting time in the pool. The advantages of this waiting time substitution have early been noticed [Irastorza & Deane 1974]. As long as an order is waiting for its release it is generally just paperwork with no material attached to it. During that time there is more **flexibility to deal with changes or cancellations**. However, many simulation studies (e.g. [Land & Gaalman 1998] in case of LOOR) have shown pool times that exceed the waiting time reductions on the floor. Though shop floor lead times decrease, total lead times increase relative to the situation without controlled release.

Generally, the poor lead time results of controlled release methods in simulation studies indicate a **lack of load balancing qualities**. It has been previously suggested that increased idle time or decreasing output would indicate a lack of balancing qualities for a (workload reducing) release method. But steady-state simulations generally use a predetermined (stochastic) order arrival process, which requires a certain long-term utilisation level to be realised. The steady state does not allow increased idle time. Instead, a larger pool of orders (see figure B.1), with an increased choice of orders to fit within the norms, must compensate for a lack of balancing qualities. As a consequence not idle time but the average pool time of orders increases in these simulations. The better the balancing qualities of the release method the smaller pool times can be. The balancing qualities of the release methods within WLC concepts are assessed in more detail in [Land & Gaalman 1998]. Analytic research of Wein [Wein 1990, Wein & Chevalier 1992] shows that an optimal release policy must give priority to improvement of the workload balance, when the orders in the pool show a small difference in urgency.

Another point brought forward in [Land & Gaalman 1998] is **that certain load fluctuations are due to the nature of job shop production, which is neglected within WLC concepts**. Job shop will generally operate on utilisation levels far below 100%. An utilisation level of 90% means that no direct load is available at a work centre for 10% of the time. It is argued that bringing the loads continuously back from zero to the norm level may lead to superfluous load under these circumstances.

B.3.3 LOOR-specific aspects

More particular strengths and weaknesses relate to the load calculation methods used within the release method. The method of LOOR estimates new inputs to the direct load by looking at the upstream distance of each order. This is shown to be particularly **important in shops with a high routing variety** [Oosterman et al. 2000]. The use of aggregate load norms neglects this routing variety and results in a poorer performance. On the contrary, the same research shows that **LOOR causes problems in shops with a dominant flow direction**, shops having typical upstream and downstream centres. There, undesirable effects are caused by the use of norms for the converted load of a typical downstream work centre. Cyclic behaviour occurs, with periods of overload alternated with periods of underload. This behaviour is related to the fact observed in [Knolmayer 1991] that **LOOR neglects the order influences after the planning period**.

Although norms on aggregate loads perform better in case of a dominant flow, the determination of adequate norm values is more difficult. The routings of orders must be considered, as typical downstream work centres require larger norms than upstream stations. **A norm for the (estimated) direct load of a work centre, as in LOOR, can more easily be related to the work centre lead-time allowance.** The development of norms for adjusted aggregate loads has led to a more robust method that works well for both shops with dominant flows and for high routing variety, while the norm values do not depend on order routings [Oosterman et al. 2000].

More particularly, points of criticism have been evoked by the assumptions of load conversion as a method for estimating inputs. A fundamental assumption is that each load account will reach its workload norm, as not the recorded level of the workload account but its norm level LPG (equation B.1 –B.4) is used to estimate input and output probabilities. Experiments in [Perona & Portioli 1996] indicate that the performance of LOOR can be improved by correcting for load deviations due to mix imbalances. In general when loose norms are applied, the norm level is not always realised, so the load conversion method will underestimate the probability that an operation can be completed during the imminent planning period. The results in [Oosterman et al. 2000] confirm that LOOR shows its best performance at tight norm levels, while **the performance at loose norm levels is relatively weak.**

Several researchers [e.g. Adam 1988, Knolmayer 1991] have criticised the input estimate from the perspective of a single order. Certainly, **other factors than the load level influence the output probability of a single order**, for instance its priority and its processing time. These are not considered in equation (B.1). Alternatively, the fact that complexity is avoided can be seen as one of the strengths of LOOR. Though for a single order, the estimation may not be realistic, the aggregated estimation of the inputs to a work centre may be close to reality. With load conversion each released order contributes a little to the converted load (see figure B.2). If the estimation incorporated the processing time of upstream orders either fully or not, one would risk larger estimate deviations depending on whether a single big order does or does not arrive according to plan. Input estimations made by LOOR will never be completely right, but deviations are generally small. Thus, **load estimations of LOOR can be qualified as simple but reasonable.**

Still, it can be argued that the load conversion method is not completely consistent with the objectives of WLC [Land & Gaalman 1996a]. This based on the fact that the estimating quality of the load conversion factor (equation B.4) improves when the operation processing times are small relative to the workload, reaching the highest accurateness for infinitely small orders. But it is the objective of WLC to reduce the load. Thus the relative lumpiness of operation times increases, and estimations may get poorer the more the general objectives are reached.

Some of the problems suggested in this section have been solved in practice by simple extension of LOOR. The next section will discuss these extensions.

B.4 LOOR – extensions based on experiences from practice

Regarding the strengths suggested in the preceding section, various experiences from practice confirm that LOOR is a powerful and easy applicable control method. User reports from industry show that lead times and work-in-progress can be reduced by up to 60% after the introduction of LOOR [Wiendahl et al. 1992, Wiendahl 1991].

However, practical experiences have indicated the relevance of individual adjustments and procedural extensions in some cases, in order to compensate for alleged weaknesses of LOOR or to consider factory-related circumstances. In this section some important extensions will be mentioned. Detailed information is provided regarding parameter setting for LOOR. The parameter determination, based on logistic operating curves, is explained by means of a case study within a job-shop of a printed-circuit-board producer [Wiendahl et al. 1998].

B.4.1 Determination of appropriate parameter values

Subsection 3.1 mentioned possible problems regarding the determination of appropriate values for control parameters and especially for the loading percentage (LPG), as performance is particularly sensitive to norm settings at low WIP-levels. In practical applications of LOOR the parameter adjustments have often been made by trial-and-error, decreasing the LPG until utilisation losses get significant [Wiendahl 1991, 1993]. Since strongly varying the parameters leads to unplanned losses in utilisation, such a procedure has to be carried out very carefully [Wiendahl 1995].

A simple method has been developed due to requirements of industry to support parameter setting in practice. This method derives the relationship between performance and WIP-levels for particular work centres from its capacity and the operation processing times. With just this information one can construct the complete logistic operating curve as depicted in figure B.6 [Wiendahl & Nyhuis 1996].

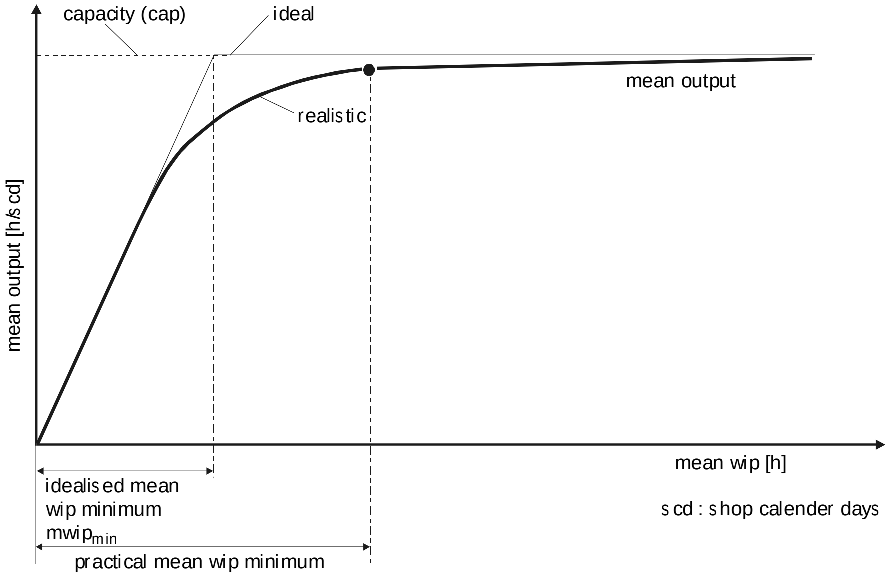

Figure B.6: Interdependency between output and work-in-process (WIP) [Nyhuis 1991]

In figure B.6 the idealised mean WIP minimum ($WIP_{m,min}$) represents the WIP level that is necessary to run the system under idealised conditions, assuming that no arriving order has to wait and no interruptions in the material flow occur. The value of the idealised mean WIP minimum can be derived directly from the operation processing times [e.g. Nyhuis & Wiendahl 1999). The realistic curve differs from the idealised one. By means of the equations (B.5) and (B.6), developed by Nyhuis, the realistic curve can be estimated for most job shop environments.

(B.5) $$WIP_m(t) = WIP_{m,\min} \cdot \left(1 - \left(1 - \sqrt[4]{t}\right)^4\right) + WIP_{m,\min} \cdot \alpha_1 \cdot t$$

(B.6) $$OUT_m(t) = OUT_{\max} \cdot \left(1 - \left(1 - \sqrt[4]{t}\right)^4\right)$$

with

$WIP_m(t)$	: mean work-in-process [h]
$OUT_m(t)$	: mean output per shop calendar day [h/scd]
$WIP_{m,min}$	: idealised mean work-in-process minimum [h]
OUT_{max}	: maximally available output [h/scd]
α_1	: stretching parameter [-]
T	: running parameter ($0 < t < 1$)

A detailed derivation of these formulas is specified in [Nyhuis 1991, Nyhuis & Wiendahl 1999]. With the aid of equation (B.5) and (B.6), a pair of values for WIP and output can be calculated dependent on the running parameter t. With these pairs, the course of the output is defined. The only parameter to specify is the α_1. Empirical research [Burmeister 1997, Nyhuis & Wiendahl 1999] and various case studies in aircraft industry, in the field of mechanical engineering and electronics industry [e.g. Wiendahl et. al. 1998] have pointed out that an α_1–value of 10 is appropriate for a wide spectrum of job shop environments. A rough sketch of the applicability of α_1–values (figure B.7) is given in [Nyhuis & Wiendahl 1999], based on data gathered in the cases studies. It is shown that only situations where a high short-term capacity flexibility strongly compensates for the loading deviations are better approached by output curves based on an α_1–value below 10. A higher α_1–value should only be considered in case of strongly fluctuating capacity requirements, resulting in high load deviations, while no short-term capacity flexibility is available to absorb these fluctuations. For a more detailed discussion the reader is referred to [Nyhuis & Wiendahl 1999].

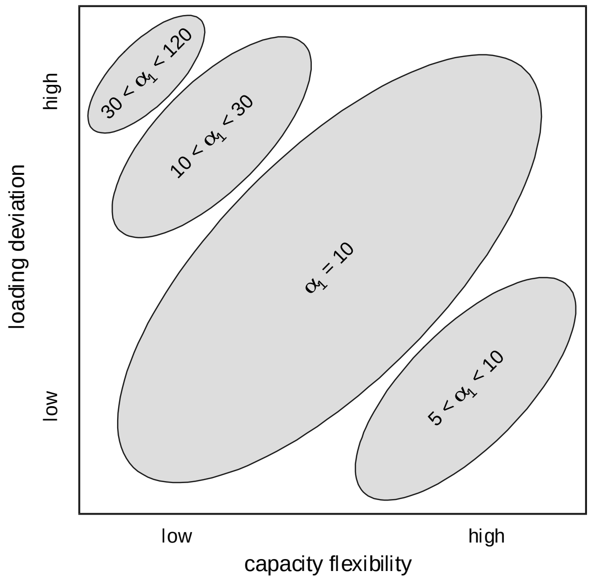

Figure B.7: Influence of loading deviations and capacity flexibility on the stretching parameter α_1 [Nyhuis & Wiendahl 1999]

Using equation (B.7), the course of the ‘range’ (which can be translated into the mean lead time) can be determined. The range can be seen as the average run-out-time of the mean work-in-process, and can be translated into the mean lead time [Nyhuis 1991].

(B.7) $R_m = \dfrac{WIP_m}{OUT_m}$

with

WIP_m : mean work-in-process [h]

OUT_m : mean output [h/scd]

R_m : mean range [scd]

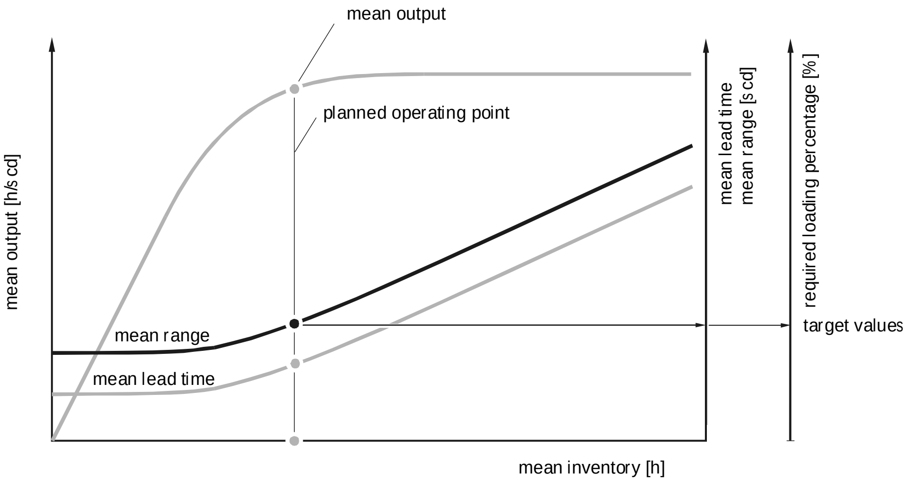

Figure B.8: Interdependency between output, lead time and work-in-process (WIP) [Nyhuis 1991]

The logistic operating curves allow to specify an appropriate operation point depending on the momentary general conditions. From that, target values for the mean range and the mean lead time can be determined easily [Nyhuis & Wiendahl 1999]. Afterwards, these values can be converted into the required loading percentage by means of equation (B.1) (Fig.8). The chosen operation point can be compared with the measured actual operation point.

This can be illustrated by a case study carried out by the Institute of Production Systems within a job shop of a printed-circuit-board-producer (PCB-producer). Within the job shop product groups such as inboard layers, non through-contacted and through-contacted printed circuit boards and multilayers are produced. The evaluation period covers a period of 5 months within which about 4300 manufacturing orders with 65000 operations have been completed at 33 work centres.

Figure B.9 shows the calculated operating curve of the bottleneck system within the job shop mentioned above. The arrow indicates the actual operating point of the work centre. The WIP-level of the work centre is far too high. It can be reduced significantly without losses in utilisation. The appropriate operating range is determined in relation to the idealised mean WIP. A WIP-level of double or triple idealised mean WIP has appeared to be suitable for most of the work centres. In case of very expensive or bottleneck work centres the WIP-level has to be increased in order to guarantee a higher utilisation. Analogous to that, WIP can be reduced if the utilisation of a work centre is not that important e.g. if the machine is already depreciated. The applicability of the operating curves to industrial problems has been confirmed in consulting projects at the Institute of Production Systems.

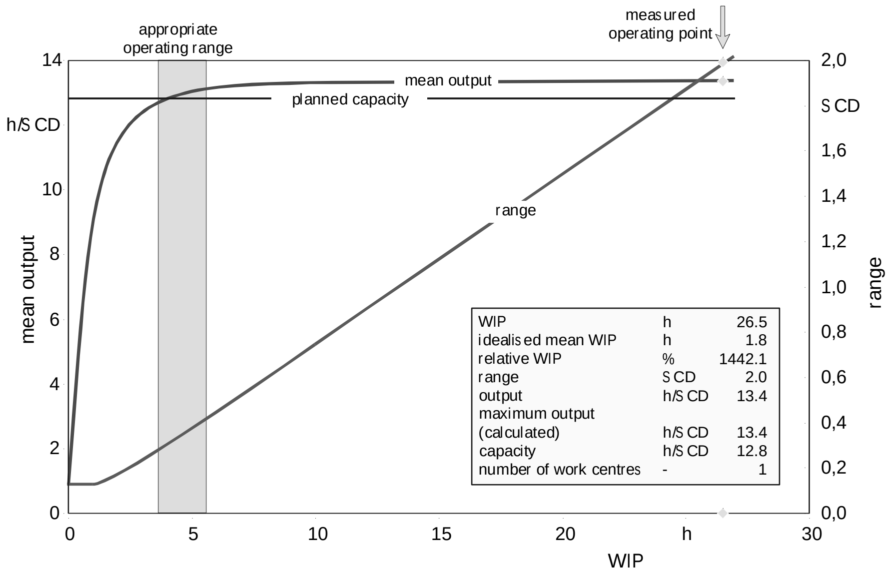

Figure B.9: Logistic operating curves of the work centre 'resistant-coating'

Adequately using operating curves can prevent LOOR from several weaknesses suggested in section 3. By means of easily determinable operating curves the problem of sensitivity to norm settings at low WIP-levels can be solved. Furthermore, it is quite easy using this technique to monitor the parameters continuously. With the aid of the operating curves the norms can be set sufficiently tight. Logistic operating curves have proven to be a powerful tool to determine appropriate values.

Another important parameter of LOOR is the anticipation horizon, which setting might be delicate (table B.1). However, investigations of the Institute of Production Systems have shown, that 2 up to 3 periods are a reasonable value in many applications in the field of mechanical engineering and electronic industry [Wiendahl 1995]. Up to now, also planning period lengths have generally been based on rules of thumb, despite the sensitivities observed in e.g. [Perona & Portioli 1998].

B.4.2 Uncertainties during the load conversion procedure

Subsection 3.2 mentioned possible balancing problems. In practice short-term capacity adjustments can often compensate for the fluctuating capacity requirements that lead to these balancing problems. These capacity adjustments can not be considered in LOOR without additional efforts. A simple change of the workload norm (e.g. by a specific change in the planned output of the following period) does not solve this problem completely, because the date of the order release does usually not correspond to the loading fluctuations at the affected centres.

In practical applications the above mentioned problem is normally solved by means of a dialogue-oriented extension of the procedure [Wiendahl 1991]. Orders not selected for release by the basic procedure discussed in section 2 are not rejected immediately. Rather, a list of work centres responsible for the rejection of orders will be created. Afterwards (e.g. in course of an agreement between sales department, production and manufacturing control) a check is made whether an order rejection is critical or if a short-term capacity adjustment is possible. This dialogue-oriented extension allows to improve load balancing and so to prevent from insufficient output. It can also be used to anticipate on possible problems in future planning periods.

Besides, this extension solves a more practical problem, as in practice the number of shop calendar days (SCD) within a planning period may vary because of holidays etc. This may cause problems as WLC methods book orders on time-independent load accounts. This is illustrated by figure B.10. Figure B.10 depicts idealised throughput diagrams of a sample work centre, based on a reduction of the planning period length from five to four shop calendar days. Each throughput diagram contains two curves, one for the cumulative input pattern at the work centre and one for the cumulative output during the planning period. Time (in shop calendar days) is set on the horizontal axis, while cumulative input and output (in hours of work) are set on the vertical axis. Realistic cumulative input and output curves would show stepwise increases, but here we use idealised curves to represent planned values. Notice that the vertical distance between the curves corresponds to the WIP level, while the horizontal distance indicates the work centre lead-time. The inclination of the output curve represents the output rate. In case (a) 40 hours of new input will be needed during the planning period to maintain the original WIP level of 30 hours until the end of the release period. This means that the work centre must be loaded up to 70 hours, which is 175% of its planned output. When the planning period counts four instead of five days, the loading percentage should be increased from 175% to 194% to maintain the original WIP level, mean lead time and output rate. In these situations, the dialogue-oriented extension allows for releasing an appropriate amount of work irrespective of the workload norms.

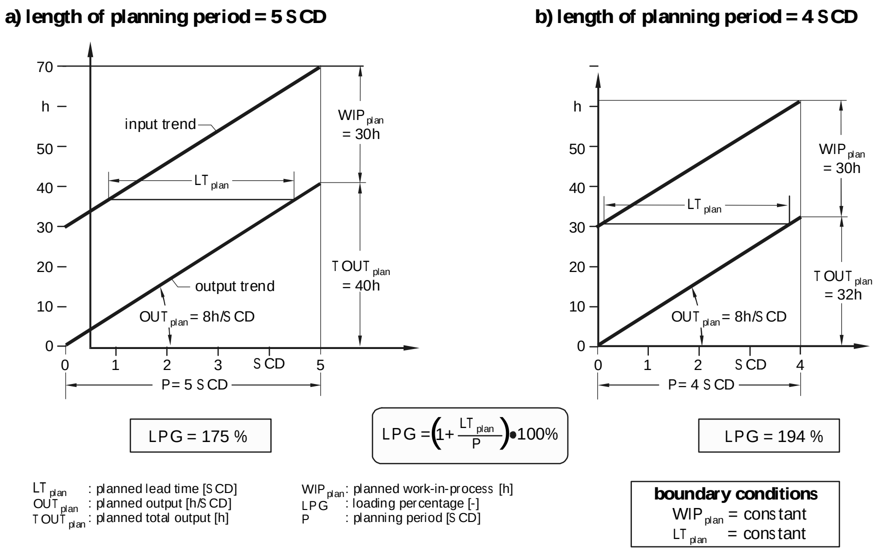

Figure B.10: Influence of planning period length on the required loading percentage

B.4.3 Rejection of orders caused by overloaded downstream work centres

Subsection 3.3 reviewed the results of simulation studies, which indicate a good performance of LOOR in typical job shops, but weaker performance in shops with a dominant flow direction. Also in practice it has been observed that unrestricted conversion of orders could initiate strong oscillations regarding the direct load for typical downstream stations. A simple but powerful solution has been found for this problem. The conversion procedure is applied only to a fixed number of process steps in advance, downstream from the current position. Thus, typical downstream operations are simply not considered when loading a job to the load accounts. Investigations of the Institute of Production Systems have confirmed that the effects mentioned above can be eliminated successfully by including only four process steps in the conversion procedure. On the one hand, this guarantees consideration of the actual loading situation at the work centres. On the other hand, it can be avoided that events which will take place in the far future influence the release of orders. This simple adjustment makes LOOR applicable in a wider range of shop configurations.

B.5 Conclusions

In the early eighties an integrated workload control concept has been developed for job shop production. Different approaches have been used to control the quantity of workload released to the shop floor, though all approaches use workload norms. The approach developed in Hannover (LOOR) has been applied in several software packages. Meanwhile, a number of researchers have assessed the release method both analytically and by means of simulation. This paper has given a review of the strengths and weaknesses of the workload control concept reported from these studies and it has discussed the extensions and adjustments that have been developed as a result of practical experiences with LOOR. The contribution of these extensions and adjustments has been related to weaknesses suggested in literature.

Table B.1 summarises the strengths and weaknesses mentioned in this paper. The strengths and weaknesses are subdivided into three categories. Category A deals with the strengths and weaknesses regarding the use of workload norms and other parameters. Category B relates to the capabilities of the WLC concepts with respect to the reduction and balancing of workloads. Category C relates to the specific strengths and weaknesses of the LOOR method.

The extensions and adjustments discussed in this paper may overcome weaknesses from all three categories. To facilitate the determination and monitoring of workload norms in practice (A), a simple method has been developed, which requires only a minimal set of data to depict the relationship between the workload level and performance for each work centre. To avoid balancing problems (B), a dialogue-oriented extension has been implemented in the practical applications LOOR. And finally, in shops with a more dominant flow direction, the performance of LOOR has been improved in shops with a dominant flow direction (C) by simply excluding typical downstream stations from release considerations.

This paper has given a rather explorative investigation of some industrial practices in relationship to alleged weaknesses of the WLC concept. Further research is required to evaluate workload control in practice more thoroughly and to evaluate the practical consequences of reported strengths and weaknesses.

Strengths	Weaknesses
A. Use of norms	
1. shop floor is buffered against disturbances 2. lead time syndromes are excluded 3. the planning interface is facilitated by norms	1. sensitive to norm setting at low WIP levels 2. anticipation horizon setting is delicate 3. sensitive to choice of planning period length 4. continuous monitoring of parameters required
B. Reducing and balancing workloads	
1. WIP is kept at a low level 2. Transparent shop without rush orders is enabled 3. No dependence on priority dispatching rules 4. lead times are made predictable 5. orders can be changed or cancelled lately	1. limited opportunities to choose efficient set-up sequences on the shop floor 2. output may drop or pool times may increase when load balancing is insufficient 3. constant norms do not consider natural load fluctuations in job shops
C. LOOR-specific	
1. job shop routing variety is considered 2. lead times and norms are easily related 3. simple but reasonable direct load estimation	1. dominant flows cause problems 2. future planning periods are neglected 3. incapable of dealing with ‘loose’ norms 4. neglects influences of priority and processing times on order progress

Table B.1: Strengths and weaknesses